WekaFS—The Data Platform That Breaks Through Storage Limits

by DCIG Lead Analyst, Storage, Ken Clipperton

WEKA

COMPANY

WekaIO, Inc.
910 E Hamilton Ave, Suite 430
Campbell, CA 95008
+1 844.392.0665
info@weka.io

www.weka.io

INDUSTRY

Data Storage

SOLUTION

WekaFS

BUSINESS USE CASES

- AI/Machine Learning
- Financial Services
- High-Performance Data Analytics
- HPC
- Life Sciences

This article discusses how Weka's high-performance software-defined data platform measures up to the DCIG's "Three New Rules for Selecting Scalable Storage Solutions." In short, WekaFS satisfies the most extreme storage requirements. It combines sophisticated, yet simple to manage, storage with a robust partnering strategy. This combination creates value for end-user organizations through quicker time to market, faster insights, and dramatic data center efficiency improvements.

Many organizations need to generate, process, and store multiple petabytes (PBs) of data. Increasingly they discover that their legacy storage solutions lack the capacity and performance required for workloads such as recommendation engines, conversational AI, risk mitigation, financial compliance, and other high-velocity analytics. Some of these workloads involve millions of tiny files, while others process individual files larger than twenty terabytes in size and individual datasets in the hundreds of terabytes.

These demands put enterprises on notice to look for and obtain scalable storage solutions that accomplish three objectives to include:

- Easily and economically grow to manage multiple petabytes of data
- Meet changing application capacity and performance demands
- Dynamically detect and adapt to changes in the environment with minimal intervention

Three New Rules for Addressing These Objectives

To select a storage solution that scales to meet these business and technical objectives, organizations should follow these three rules:

1. Select scalable storage solutions that offer scale-out architectures
2. Select scalable storage solutions that offer both NAS and S3 object storage interfaces
3. Select scalable storage solutions that natively offer connectivity to third-party storage clouds

Rule #1: Select scalable storage solutions that offer scale-out architectures

WekaFS is a modern file system designed on day one as scale-out architecture with data structures optimized for distributed grids and the parallelism of NVMe-flash. Many file operations involve only metadata. By keeping all metadata in NVMe flash and

always writing new data to NVMe-flash, Weka delivers latencies of just 200-300 microseconds, even as capacity scales to hundreds of petabytes.

The Weka software runs in a Linux container on industry-standard x86 server hardware. Nodes can run as dedicated storage or as converged compute+storage infrastructure. Clusters can start small yet scale to hundreds of nodes from multiple server generations and multiple vendors. This deployment flexibility enables organizations to dynamically and independently scale cluster capacity and performance up or down to meet extreme storage demands—economically.

> *"Weka's record-breaking performance and robust partnering strategy translate into real business value for Weka customers."*
>
> **— Ken Clipperton, DCIG Lead Analyst, Storage**

Weka focuses more on partnering than competing, creating complete pre-integrated high-performance solutions rather than being yet another point solution that customers must somehow integrate into their technology infrastructures. It even partners with other storage vendors. The resulting solutions are sold by key technology partners including Cisco, Dell, HPE, Hitachi Vantara, Lenovo, Penguin Computing, Supermicro, and specialty solution providers.

Weka's record-breaking performance and robust partnering strategy translate into real business value for Weka customers. Their joint solutions enable end-user organizations to achieve quicker time to market, faster insights, and dramatic data center efficiency improvements.

Rule #2: Select scalable storage solutions that offer both NAS and S3 object storage interfaces

Organizations must re-think how they store and access their increasingly large data stores using a wider variety of applications. They must select storage solutions based on their capacity, performance, and the networked storage protocols they support. Some legacy NAS providers are responding by adding S3 object protocol support, and object storage providers are adding NAS file protocol support.

WekaFS supports POSIX, NFS, SMB, and now NVIDIA® GPUDirect® Storage (GDS) protocols. With Weka, data is fully sharable between all the protocols—no tuning required. This extensive protocol support eliminates storage silos and enables a single copy of data to serve all users and applications.

WekaFS employs GPUDirect Storage to deliver world-record-shattering performance and maximize the amount of work that NVIDIA-based systems can accomplish.

GDS provides the following technical benefits:[1]

- Enables a direct path between GPU memory and storage.
- Increases the bandwidth, reduces the latency, and reduces the load on CPUs and GPUs for data transferal.
- Reduces the performance impact and dependence on CPUs to process storage data transfer.
- Performance force multiplier on top of the compute advantage for computational pipelines that are fully migrated to the GPU so that the GPU, rather than the CPU, has the first and last touch of data that moves between storage and the GPU.
- Supports interoperability with other OS-based file access, enabling data to be transferred to and from the device by using traditional file IO.

These technical benefits translate into real business value for Weka customers. For example, Weka reduced one customer's pipelines from two weeks to four hours. Another Weka customer had a genomics workload that was only 20% complete after 70 days of processing on legacy NAS. Weka completed the same workload in just 7 days. Accelerating these workloads by 80x and 50x delivered much better infrastructure efficiency and, even more importantly, accelerated innovation through reduced time to insight.

Rule #3: Select scalable storage solutions that natively offer connectivity to third-party storage clouds

Enterprises that store petabytes of data often have an increasing need to store some or all this data off-site. They may store it with these providers for archival, backup, business continuity, compliance, bursting workloads to the cloud, or distributing access to the data. Regardless of the organization's rationale, scalable storage solutions must seamlessly store data with third-party cloud storage providers.

WekaFS more than meets these requirements. It supported native connectivity to S3 from day one. Its integrated tiering layer automatically offloads cold data to any S3 object store or cloud, applying user-defined policies on a per-filesystem basis. But Weka does more than tier data to the cloud. Its global namespace extends to data in these cloud object stores. Thus, all the data remains available to all applications.

Organizations can run Weka in the cloud, but if an organization uses the cloud only for backup and disaster recovery, it is not necessary to keep a Weka instance running continuously. Instead, customers can spin up a cloud cluster as needed to receive snapshots and then spin the cluster down. This flexibility can result in substantial dollar savings compared to solutions that must run continuously in the cloud.

Beyond providing outstanding performance and integrating with cloud storage, Weka offers a rich set of data services. These enable Weka to integrate cleanly into existing enterprise infrastructures, automation frameworks, and data protection schemes.

The Rules Have Changed

Organizations needing to store a petabyte or more of data should re-examine the storage systems they use. New scale-out storage solutions have sufficiently matured to offer the ease of deployment and management of scale-up storage systems. These scale-out solutions also better address the long-standing data migration and hardware refreshes associated with scale-up storage.

More importantly, new scale-out storage solutions offer additional features that organizations need to manage petabytes of data. They support multiple networked storage protocols, including S3, with the underlying technologies to deliver high performance levels. Further, they recognize and support the use of third-party storage clouds to meet other business needs.

This combination of factors puts organizations on notice. When it comes to supporting data stores of over a petabyte, the rules for selecting scalable storage solutions have changed.

WekaFS meets and exceeds even the most extreme storage demands. Combining its high-performance scale-out, multi-protocol, cloud-integrated file technology with a robust partnering strategy, Weka has created a solution that offers quick time-to-value through a network of trusted enterprise technology providers.

Any organization grappling with multi-petabyte data management challenges or seeking to create new value through data-intensive processing would do well to examine the Weka solution. ■

---

1. https://docs.nvidia.com/gpudirect-storage/overview-guide/

About WekaIO

WekaIO (Weka) is used by 8 of the Fortune 50 enterprise organizations to maximize the value out of their data. Purpose-built to unlock the full capabilities for the data center of today and the future, the Weka Limitless data platform eliminates the tradeoff between Simplicity, Speed and Scale. Built on WekaFS, the Limitless data platform is optimized for NVMe-flash and the hybrid cloud so that customers can capitalize on the advances in Cloud, GPU Compute and Flash technologies without any compromises. For more information, please visit **https://www.weka.io/***.*

DCIG, LLC // 7511 MADISON STREET // OMAHA NE 68127 // 844.324.4552   **dcig.com**